wwPDB NMR Structure Validation Summary Report

Jun 15, 2024 – 07:52 PM EDT

PDB ID	:	2KSC
BMRB ID	:	16306
Title	:	Solution structure of Synechococcus sp. PCC 7002 hemoglobin
Authors	:	Vuletich, D.A.; Falzone, C.J.; Lecomte, J.T.J.
Deposited on	:	2010-01-02

This is a wwPDB NMR Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

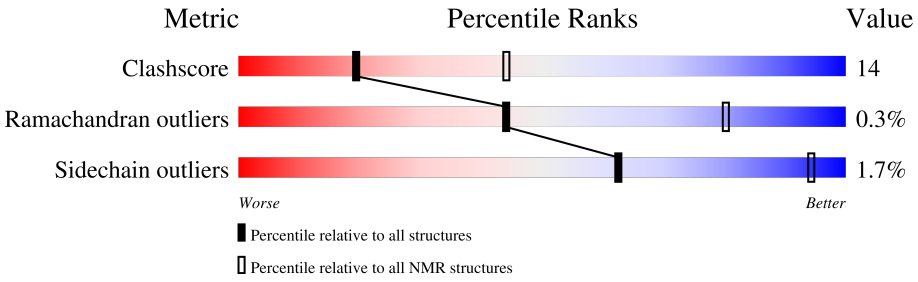
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 90%.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and a grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Mol	Chain	Length	Quality of chain
1	A	123	61% 15% • 23%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA and RNA chains that are outliers for geometric criteria:

Mol	Chain	Compound	Res	Total models with violations	
				Chirality	Geometry
2	A	HEB	125	16	-

2 Ensemble composition and analysis

This entry contains 16 models. Model 1 is the overall representative, medoid model (most similar to other models).

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:21-A:56, A:66-A:124 (95)	0.61	1

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 3 clusters and 4 single-model clusters were found.

Cluster number	Models
1	1, 4, 6, 7, 9, 11, 13
2	5, 14, 15
3	3, 8
Single-model clusters	2; 10; 12; 16

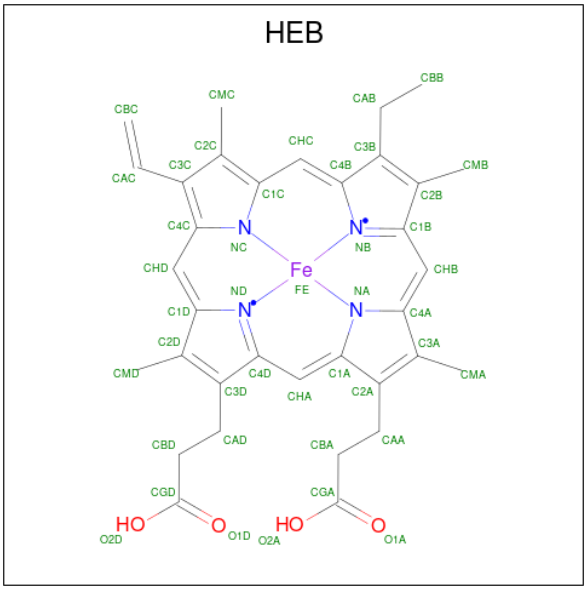
3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1990 atoms, of which 980 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Cyanoglobin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	123	1916	605	949	172	187	3	0

- Molecule 2 is HEME B/C (three-letter code: HEB) (formula: $C_{34}H_{34}FeN_4O_4$).



Mol	Chain	Residues	Atoms					
			Total	C	Fe	H	N	O
2	A	1	74	34	1	31	4	4

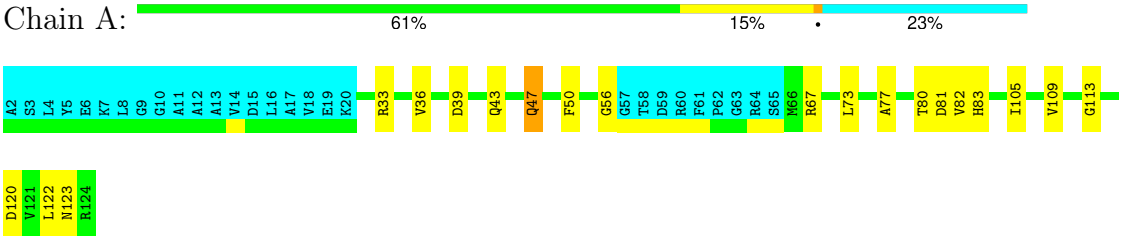
4 Residue-property plots

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

- Molecule 1: Cyanoglobin

Chain A:

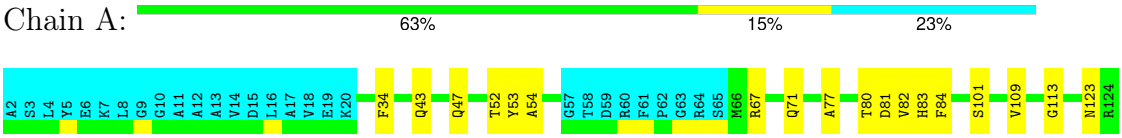


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: Cyanoglobin

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *DGSA-distance geometry simulated annealing, torsion angle dynamics.*

Of the 30 calculated structures, 16 were deposited, based on the following criterion: *structures with the least restraint violations.*

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
X-PLOR NIH	structure solution	2.23
X-PLOR NIH	refinement	2.23

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	1460
Number of shifts mapped to atoms	1429
Number of unparsed shifts	0
Number of shifts with mapping errors	31
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	90%

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: HEB

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	765	745	745	21±5
2	A	43	31	31	3±1
All	All	12928	12416	12416	358

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 14.

5 of 114 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:26:LEU:HD13	1:A:27:ALA:N	0.94	1.77	16	1
1:A:26:LEU:HD22	1:A:26:LEU:O	0.82	1.73	16	1
1:A:73:LEU:HD11	2:A:125:HEB:HMD2	0.77	1.56	6	14
1:A:81:ASP:OD2	1:A:122:LEU:HD11	0.73	1.83	11	3
1:A:54:ALA:O	1:A:111:ILE:HG21	0.73	1.83	11	4

6.3 Torsion angles (i)

6.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the backbone conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	94/123 (76%)	92±2 (98±2%)	2±1 (2±1%)	0±1 (0±1%)	44 80
All	All	1504/1968 (76%)	1469 (98%)	31 (2%)	4 (0%)	44 80

All 4 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	82	VAL	1
1	A	101	SER	1
1	A	123	ASN	1
1	A	113	GLY	1

6.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the sidechain conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	83/102 (81%)	82±1 (98±1%)	1±1 (2±1%)	62 94
All	All	1328/1632 (81%)	1306 (98%)	22 (2%)	62 94

5 of 7 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	47	GLN	11
1	A	26	LEU	4
1	A	42	LYS	2
1	A	102	GLN	2
1	A	24	LYS	1

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	HEB	A	125	1	47,50,50	1.01±0.01	3±0 (6±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	HEB	A	125	1	55,82,82	0.96±0.01	1±0 (1±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means

no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	HEB	A	125	1	1±0,1,8,8	0±0,12,54,54	-

All unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
2	A	125	HEB	C3C-C2C	3.66	1.35	1.40	16	16
2	A	125	HEB	O2A-CGA	2.57	1.22	1.30	3	16
2	A	125	HEB	O2D-CGD	2.55	1.22	1.30	1	16

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	A	125	HEB	CAB-C3B-C4B	2.82	128.46	124.79	1	16

All unique chiral outliers are listed below.

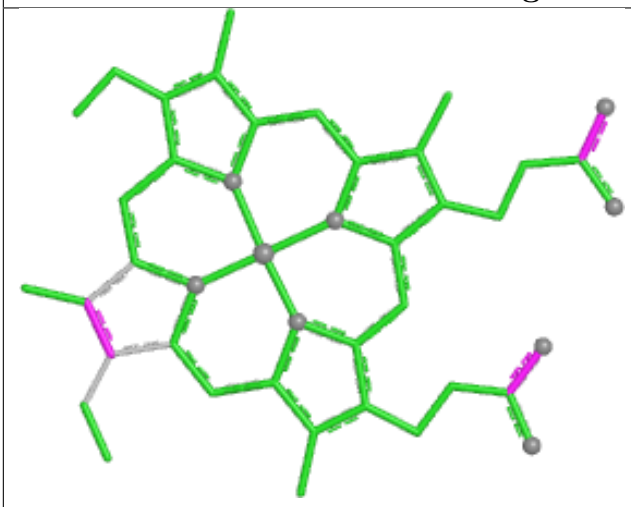
Mol	Chain	Res	Type	Atoms	Models (Total)
2	A	125	HEB	NA	16

There are no torsion outliers.

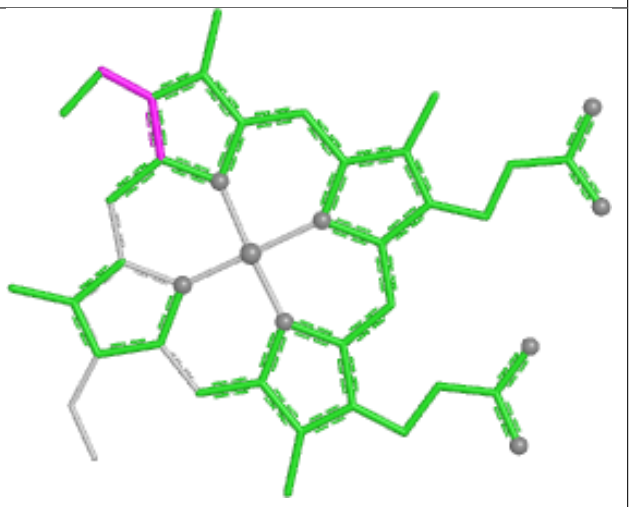
There are no ring outliers.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

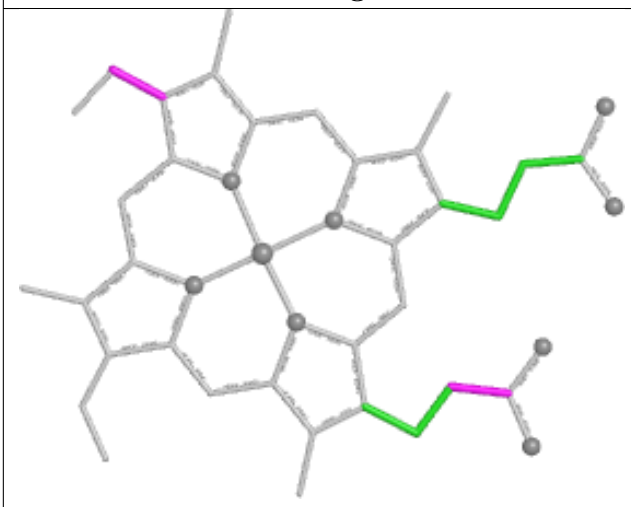
Ligand HEB A 125



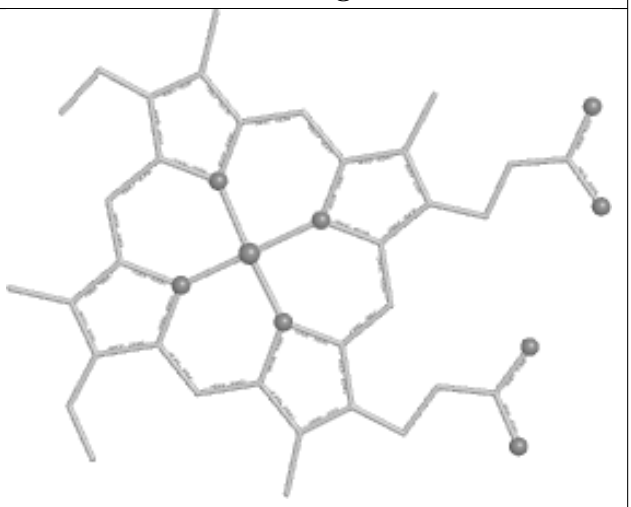
Bond lengths



Bond angles



Torsions



Rings

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 90% for the well-defined parts and 85% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	1460
Number of shifts mapped to atoms	1429
Number of unparsed shifts	0
Number of shifts with mapping errors	31
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	43

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. First 5 (of 31) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	UNMAPPED	1	HEB	HHA	-2.3	0.05	1
1	UNMAPPED	1	HEB	HHB	0.61	0.05	1
1	UNMAPPED	1	HEB	HHC	1.06	0.05	1
1	UNMAPPED	1	HEB	HHD	-1.26	0.05	1
1	UNMAPPED	1	HEB	HAB2	2.75	0.02	1
1	UNMAPPED	1	HEB	HAC	6.67	0.02	1
1	UNMAPPED	1	HEB	HAA1	3.52	0.02	2
1	UNMAPPED	1	HEB	HAA2	2.63	0.02	2
1	UNMAPPED	1	HEB	HAD1	8.4	0.02	2
1	UNMAPPED	1	HEB	HAD2	7.75	0.02	2
1	UNMAPPED	1	HEB	HBA1	-0.11	0.02	2
1	UNMAPPED	1	HEB	HBA2	-0.57	0.02	2
1	UNMAPPED	1	HEB	HBB1	0.82	0.02	1
1	UNMAPPED	1	HEB	HBB2	0.82	0.02	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	UNMAPPED	1	HEB	HBB3	0.82	0.02	1
1	UNMAPPED	1	HEB	HBC1	-2.04	0.02	1
1	UNMAPPED	1	HEB	HBC2	-2.44	0.02	1
1	UNMAPPED	1	HEB	HBD1	0.46	0.02	2
1	UNMAPPED	1	HEB	HBD2	0.75	0.02	2
1	UNMAPPED	1	HEB	HMA1	13.27	0.02	1
1	UNMAPPED	1	HEB	HMA2	13.27	0.02	1
1	UNMAPPED	1	HEB	HMA3	13.27	0.02	1
1	UNMAPPED	1	HEB	HMB1	11.0	0.02	1
1	UNMAPPED	1	HEB	HMB2	11.0	0.02	1
1	UNMAPPED	1	HEB	HMB3	11.0	0.02	1
1	UNMAPPED	1	HEB	HMC1	12.05	0.02	1
1	UNMAPPED	1	HEB	HMC2	12.05	0.02	1
1	UNMAPPED	1	HEB	HMC3	12.05	0.02	1
1	UNMAPPED	1	HEB	HMD1	19.41	0.02	1
1	UNMAPPED	1	HEB	HMD2	19.41	0.02	1
1	UNMAPPED	1	HEB	HMD3	19.41	0.02	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	116	-0.99 ± 0.17	Should be checked
$^{13}C_{\beta}$	109	0.28 ± 0.09	None needed (< 0.5 ppm)
$^{13}C'$	115	-0.70 ± 0.12	Should be applied
^{15}N	118	0.15 ± 0.19	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 90%, i.e. 1195 atoms were assigned a chemical shift out of a possible 1327. 0 out of 21 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	479/479 (100%)	194/194 (100%)	190/190 (100%)	95/95 (100%)
Sidechain	631/742 (85%)	433/480 (90%)	180/229 (79%)	18/33 (55%)
Aromatic	85/106 (80%)	44/54 (81%)	39/48 (81%)	2/4 (50%)
Overall	1195/1327 (90%)	671/728 (92%)	409/467 (88%)	115/132 (87%)

7.1.4 Statistically unusual chemical shifts (i)

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	46	HIS	HB2	11.84	1.36 – 4.85	25.0
1	A	70	HIS	HB2	9.92	1.36 – 4.85	19.5
1	A	70	HIS	HB3	9.45	1.18 – 4.91	17.2
1	A	46	HIS	HB3	9.20	1.18 – 4.91	16.5
1	A	121	VAL	HG11	4.15	-0.48 – 2.12	12.8
1	A	121	VAL	HG12	4.15	-0.48 – 2.12	12.8
1	A	121	VAL	HG13	4.15	-0.48 – 2.12	12.8
1	A	46	HIS	HA	9.41	2.49 – 6.71	11.4
1	A	46	HIS	CA	82.50	45.04 – 67.94	11.4
1	A	121	VAL	HG21	3.78	-0.58 – 2.19	10.7
1	A	121	VAL	HG22	3.78	-0.58 – 2.19	10.7
1	A	121	VAL	HG23	3.78	-0.58 – 2.19	10.7
1	A	50	PHE	HD2	3.76	5.52 – 8.61	-10.7
1	A	50	PHE	HD1	3.76	5.51 – 8.60	-10.7
1	A	69	ALA	HB1	-0.81	0.14 – 2.58	-8.9
1	A	69	ALA	HB2	-0.81	0.14 – 2.58	-8.9
1	A	69	ALA	HB3	-0.81	0.14 – 2.58	-8.9
1	A	50	PHE	HE1	4.54	5.56 – 8.62	-8.3
1	A	50	PHE	HE2	4.54	5.54 – 8.63	-8.2
1	A	87	ILE	HD11	-1.23	-0.72 – 2.09	-6.8
1	A	87	ILE	HD12	-1.23	-0.72 – 2.09	-6.8
1	A	87	ILE	HD13	-1.23	-0.72 – 2.09	-6.8
1	A	87	ILE	HG21	-1.01	-0.56 – 2.11	-6.7
1	A	87	ILE	HG22	-1.01	-0.56 – 2.11	-6.7
1	A	87	ILE	HG23	-1.01	-0.56 – 2.11	-6.7
1	A	70	HIS	CA	71.70	45.04 – 67.94	6.6
1	A	47	GLN	HA	6.96	2.17 – 6.35	6.5
1	A	66	MET	HB3	4.09	0.33 – 3.66	6.3
1	A	121	VAL	HA	7.49	1.33 – 6.99	5.9
1	A	73	LEU	HD11	-0.80	-0.61 – 2.12	-5.7
1	A	73	LEU	HD12	-0.80	-0.61 – 2.12	-5.7
1	A	73	LEU	HD13	-0.80	-0.61 – 2.12	-5.7
1	A	33	ARG	HG3	0.02	0.15 – 2.94	-5.5
1	A	87	ILE	HB	0.23	0.35 – 3.22	-5.4
1	A	79	LEU	HD11	-0.69	-0.61 – 2.12	-5.3
1	A	79	LEU	HD12	-0.69	-0.61 – 2.12	-5.3
1	A	79	LEU	HD13	-0.69	-0.61 – 2.12	-5.3
1	A	47	GLN	HB2	3.35	0.80 – 3.29	5.2

Continued on next page...

Continued from previous page...

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	121	VAL	HB	3.61	0.43 – 3.54	5.2
1	A	66	MET	HE1	-0.09	-0.03 – 3.80	-5.2
1	A	66	MET	HE2	-0.09	-0.03 – 3.80	-5.2
1	A	66	MET	HE3	-0.09	-0.03 – 3.80	-5.2
1	A	66	MET	HB2	3.66	0.42 – 3.63	5.1

7.1.5 Random Coil Index (RCI) plots

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: